Efficient Streaming of Stereoscopic Depth-Based
3D Videos

D. Temel, M. Aabed, M. Solh and G. AlRegib
School of Electrical and Computer Engineering
Georgia Tech, Atlanta, GA
Outline

1. Description
2. Methodology
   I. Depth-Map Cues Extraction
   II. Luminance/Chrominance Based Depth Extraction
   III. Texture Based Depth Extraction
   IV. Motion Based Depth Extraction
   V. Combining Motion and Texture Based Depths
3. Simulation results
4. Conclusion
Luminance/Chrominance Based Depth Extraction

Depth cues

Luminance Based Depth Extraction

Y based depth estimation

U based depth estimation

V based depth estimation

Chrominance Based Depth Extraction

\[ \overline{Z_{CR}} = C_{RI} \times \frac{(D_{N-1} - D_0)}{255} + D_0 \]

\[ \overline{Z_{CB}} = C_{BI} \times \frac{(D_{N-1} - D_0)}{255} + D_0 \]
Texture Based Depth Extraction

Reference Image → Texture Cues Extraction → Depth Map Estimation → Estimated Depth Map

\[ t[x, y] = EI[x, y] \times MI[x, y] \]

\[ T[x, y] = \begin{cases} 
K_1 + \left( 0.5 \times K_1 \times \frac{\log_2 t[x, y]}{\log_2 \beta_1} \right) & \beta_1 \leq t[x, y] < \beta_2 \\
K_2 + \left( 0.5 \times K_2 \times 2^{-(t[x, y] - \beta_2)} \right) & t[x, y] > \beta_2 \\
K_1 & \text{otherwise}
\end{cases} \]

\[ Z_T = \frac{IT - \min (IT)}{\min (IT) - \max (IT)} \times (D_{N-1} - D_0) + D_0 \]
Motion Based Depth Extraction

Reference Images → Motion Cues Extraction → Depth Map Estimation → Motion Based Depth Estimation

\[ Z_M = \frac{IM - \min(IM)}{\min(IM) - \max(IM_i)} \times (D_{N-1} - D_0) + D_0 \]
Combining Motion and Texture Based Depths

Motion based estimation

Texture based estimation

Motion-Texture Combination

Ground Truth

\[ S_1[m,n] = \begin{cases} 
Z_M[m,n], & Z_M[m,n] \neq 0 \text{ and } Z_T[m,n] = 0 \\
Z_T[m,n], & Z_M[m,n] = 0 \text{ and } Z_T[m,n] \neq 0 \\
A_1[m,n], & Z_M[m,n] = 0 \text{ and } Z_T[m,n] = 0 \\
sZ_M[m,n] + (1 - s)Z_T[m,n], & \text{otherwise}
\end{cases} \]

\[ A_{2M}[m,n] = \begin{cases} 
Z_M[m,n+k], & Z_M[m,n+k] > \epsilon \\
0, & \text{otherwise}
\end{cases} \]

\[ A_1[m,n] = \begin{cases} 
A_{2M}[m,n], & A_{2M}[m,n] \neq 0 \text{ and } A_{2T}[m,n] = 0 \\
A_{2T}[m,n], & A_{2M}[m,n] = 0 \text{ and } A_{2T}[m,n] \neq 0 \\
sA_{2M} + (1 - s)A_{2T}[m,n], & A_{2M}[m,n] \neq 0 \text{ and } A_{2T}[m,n] \neq 0
\end{cases} \]
Combining Motion and Texture Based Depths

Motion-Texture Combination

Ground Truth
Simulation Results: \(3VQM\) and \(PSNR\)

<table>
<thead>
<tr>
<th>Sequence</th>
<th>(3VQM)</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M-T</td>
<td>Y</td>
</tr>
<tr>
<td>Balloons</td>
<td>4.24</td>
<td>2.16</td>
</tr>
<tr>
<td>Cafe</td>
<td>4.69</td>
<td>3.03</td>
</tr>
<tr>
<td>Champ. Tower</td>
<td>4.07</td>
<td>0.41</td>
</tr>
<tr>
<td>Kendo</td>
<td>4.21</td>
<td>2.90</td>
</tr>
<tr>
<td>Love Birds</td>
<td>4.84</td>
<td>3.98</td>
</tr>
<tr>
<td>Pantomime</td>
<td>4.44</td>
<td>0.17</td>
</tr>
</tbody>
</table>

\[
3VQM = K(1 - SO(SO \cap TO))a(1 - TI)b(1 - TO)c
\]

**Spatial Outliers**: Standard deviation of depth map errors

\[SO = STD(\Delta Z)\]

**Temporal Outliers**: Standard deviation of two depth map errors in time domain

\[TO = STD(\Delta Z_{t+1} - \Delta Z_t)\]

**Temporal Inconsistencies**: Standard deviation of two depth values in two time instances

\[TI = STD(Z_{t+1} - Z_t)\]

\[\Delta Z = |Z_{IDEAL} - Z|\]
Simulation Results:
Pantomime depth maps and synthesized views

Ground-truth depth
Depth from Y
Depth from U
Depth from V

Synthesized using GT
Synthesized using Y
Synthesized using U
Synthesized using V
Simulation Results: Streaming

Efficient Streaming of Stereoscopic Depth-Based 3D Videos

<table>
<thead>
<tr>
<th>Image Sequence</th>
<th>Compressed(2D + Depth map)(Bytes)</th>
<th>Compressed(2D) + Depth cues(Bytes)</th>
<th>Savings ratio(“%”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balloons</td>
<td>254,648</td>
<td>155,686</td>
<td>38.86</td>
</tr>
<tr>
<td>Champ. Tower</td>
<td>240,189</td>
<td>161,478</td>
<td>32.77</td>
</tr>
<tr>
<td>Kendo</td>
<td>356,271</td>
<td>216,846</td>
<td>39.13</td>
</tr>
<tr>
<td>Love Birds</td>
<td>156,410</td>
<td>129,143</td>
<td>17.43</td>
</tr>
<tr>
<td>Pantomime</td>
<td>630,024</td>
<td>536,622</td>
<td>14.83</td>
</tr>
</tbody>
</table>
Questions

For more information and related work:
http://www.ece.gatech.edu/research/labs/msl/
Simulation Results: Balloons
Simulation Results: Cafe

Depth from Y  Depth from U  Depth from V  Depth from MT  Ground-truth depth

Synthesized using Y  Synthesized using U  Synthesized using V  Synthesized using M-T  Synthesized using GT
Simulation Results:
Champagne Tower

Depth from Y
Depth from U
Depth from V
Depth from MT
Ground-truth depth

Synthesized using Y
Synthesized using U
Synthesized using V
Synthesized using M-T
Synthesized using GT
Simulation Results: Kendo

Depth from Y, Depth from U, Depth from V, Depth from MT, Ground-truth depth

Synthesized using Y, Synthesized using U, Synthesized using V, Synthesized using M-T, Synthesized using GT
Simulation Results:
*Love Birds*

- Depth from Y
- Depth from U
- Depth from V
- Depth from MT
- Ground-truth depth

- Synthesized using Y
- Synthesized using U
- Synthesized using V
- Synthesized using M-T
- Synthesized using GT